SUPPLEMENT
The uncertainty principle of measurement in vision

Sergei Gepshtein[1,2], Ivan Tyukin[3], and Thomas D. Albright[1]

[1]The Salk Institute, USA; [2]RIKEN BSI, Japan; [3]Leicester University, UK

July 2, 2010

Abstract

We review the reasoning underlying two approaches to combination of sensory uncertainties. First approach is noncommittal, making no assumptions about properties of uncertainty or parameters of stimulation, as in Gepshtein, Tyukin, and Albright (2010). Then we explain the relationship between this approach and the one commonly used in modeling "higher level" aspects of sensory systems, such as in visual cue integration, where assumptions are made about properties of stimulation. The two approaches follow similar logic, except in one case maximal UNCERTAINTY is minimized and in the other minimal CERTAINTY is maximized. Then we demonstrate how optimal solutions are found to the problem of resource allocation under uncertainty.

Contents

1. Combination of uncertainties

1.1. Noncommittal approach

Let the stimulus be an integrable function of one variable $I(x)$ that depends on two aspects of stimulation:

- Stimulus location on x, where x can be space or time, the "location" indicating, respectively *where* or *when* stimulation occurred.
- Stimulus content on f_x, where f_x can be spatial or temporal frequency of stimulus modulation.

We consider a sensory system equipped with many measuring devices, each able to estimate both stimulus location and content from $I(x)$. We assume that the error of estimation is a random variable with probability density $p(x, f)$.

It is sometimes assumed that sensory systems know $p(x, f)$: a case we review in the next section. But in general we do not know $p(x, f)$; we only know (or guess) some of its properties, such as its mean value and variance. In particular, let

$$\begin{aligned} p_x(x) &= \int p(x, f)df, \\ p_f(f) &= \int p(x, f)dx \end{aligned} \tag{S1}$$

be the (marginal) means of $p(x, f)$ on dimensions x and f_x. Sensory systems can optimize their performance with this minimal knowledge, as follows.

To reduce the chances of making gross errors, we use the following strategy. We find the condition of *minimal* uncertainty against the profile of *maximal* uncertainty, i.e., using a minimax approach (von Neumann, 1928; Luce & Raiffa, 1957). We do so in two steps. First we find such $p_x(x)$ and $p_f(f)$ for which measurement uncertainty is maximal. Then we find the condition at which the function of maximal uncertainty has the smallest value: the minimax point.

We evaluate maximal uncertainty using the well-established definition of entropy (Shannon, 1948):

$$H(X, F) = -\int p(x, f) \log p(x, f)dx\, df.$$

Recall that Shannon's entropy is sub-additive:

$$H(X, F) \leq H(X) + H(F) = H^*(X, Y), \tag{S2}$$

where

$$\begin{aligned} H(X) &= -\int p_x(x) \log p_x(x)dx, \\ H(F) &= -\int p_f(f) \log p_f(f)df. \end{aligned}$$

Therefore, we can say that the uncertainty of measurement cannot exceed

$$\begin{aligned} H^*(X, F) = &-\int p_x(x) \log p_x(x)dx \\ &-\int p_f(f) \log p_f(f)df. \end{aligned} \tag{S3}$$

Eq. S3 is the "envelope" of maximal measurement uncertainty: a "worst-case" estimate.

By the Boltzmann theorem on maximum-entropy probability distributions (Cover & Thomas, 2006), the maximal entropy of probability densities with fixed means and variances is attained when the functions are Gaussian. Then, maximal entropy is a sum of their variances (Cover & Thomas, 2006). We obtain

$$p_x(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-x^2/2\sigma_x^2},$$
$$p_f(f) = \frac{1}{\sigma_f\sqrt{2\pi}} e^{-f^2/2\sigma_f^2},$$

where σ_x and σ_f are the standard deviations. And the maximal entropy is simply:

$$H = \sigma_x^2 + \sigma_f^2. \tag{S4}$$

That is, when variances are unknown, maximal uncertainty of measurement is a sum of variances of measurement components.

This is the method used by Gepshtein et al. (2007, 2010) in derivations of *joint uncertainty* and *composite uncertainty* functions.[1] The authors then found the optimal conditions by looking for minimal values of the uncertainty functions.

1.2. Top-down approach

Now we assume the system enjoys some knowledge of stimulation, so we can use *likelihood* as a measure of uncertainty. Suppose we want to derive a combined estimate z from two estimates x and f of some parameter of stimulation. We assume that likelihood functions $P(z|x, f)$, $P_x(z|x)$, and $P_f(z|f)$ are continuous, differentiable, and known. Let us first assume that likelihoods are separable:

$$P(z|x, f) = P_x(z|x)P_f(z|f). \tag{S5}$$

Then, the most likely value of z is

$$z^* = \arg\max_z P(z|x, f) = \arg\max_z[\log P_x(z|x) + \log P_f(z|f)].$$

We can use the logarithmic transformation because it is a strictly monotone continuous function on $(0, \infty)$, and hence it does not change maxima of continuous functions.

It is commonly assumed that $P_x(z|x)$ and $P_f(z|f)$ are Gaussian functions, or that they are well approximated by Gaussian functions. For example, Yuille and Bülthoff (1996) assumed that cubic and higher-order terms of the Taylor expansion of $\log P_x(z|x)$ can be neglected, which is equivalent to the assumption of Gaussianity. (We return to this assumption, and also the assumption of separability in a moment.) Then

$$P_x(z|x) = c_x e^{-(z-z_x)^2/2\sigma_x^2},$$
$$P_f(z|f) = c_f e^{-(z-z_f)^2/2\sigma_f^2}, \ c_x, c_f \in \mathbb{R}_{>0}$$

and

$$\log P_x(z|x) + \log P_f(z|f) =$$
$$\log c_x + \log c_f - \frac{1}{2\sigma_x^2}(z - z_x)^2 - \frac{1}{2\sigma_f^2}(z - z_f)^2.$$

[1] For simplicity, Gepshtein et al. (2010) use intervals of measurement, rather than interval variances, as estimates of component uncertainties.

The latter expression is maximized when its first derivative over z is zero. Hence

$$
\begin{aligned}
z^* &= \left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_f^2}\right)^{-1} \left(\frac{1}{\sigma_x^2} z_x + \frac{1}{\sigma_f^2} z_f\right) \\
&= \frac{1}{\sigma_x^2 + \sigma_f^2} \left(\sigma_f^2 z_x + \sigma_x^2 z_f\right),
\end{aligned}
\tag{S6}
$$

which is the familiar weighted-average rule of cue combination (Cochran, 1937; Maloney & Landy, 1989; Clark & Yuille, 1990; Landy, Maloney, Johnsten, & Young, 1995; Yuille & Bülthoff, 1996). In general, when the number of measurements is greater than two, the combination rule of Eq. S6 becomes

$$
z^* = \frac{1}{\sum_i \sigma_i^2} \sum_i z_i \prod_{j \neq i} \sigma_j^2, \tag{S7}
$$

where z_i are such that individual likelihood functions attain their maxima at z_i.

Why is the assumption common that likelihood functions have the simple form of Eq. S5, i.e., are separable and Gaussian? An answer follows from the argument we presented in the previous section. Suppose that one seeks to estimate the likelihood function when its shape is unknown. We saw in the previous section that the *least certain* estimate is the likelihood function for which the entropy is maximal. Hence, by sub-additivity of entropy (Eq. S2), the least certain estimate of $P(z|x, f)$ is

$$
P(z|x, f) = P_x(z|x) P_f(z|f),
$$

as in Eq. S5. Moreover, if the mean values and variances of $P_x(z|x)$ and $P_f(z|f)$ are fixed, then the likelihood functions must be Gaussian, by the same argument. Indeed, separable Gaussian likelihood functions are the least certain estimates.

2. Resource allocation

In Gepshtein et al. (2010) we asked how sensory system ought to allocate their resources in face of uncertainties inherent in measurement and stimulation. We approached this problem in two steps. First, we combined all uncertainties in *uncertainty functions*: comprehensive descriptions of how quality of measurement varied across conditions of measurement. Second, we proposed how limited resources are to be allocated given the uncertainty functions. Here we illustrate the second step in more detail, using the approach of constrained optimization.

A key requirement of allocation is to optimize reliability (reduce uncertainty) of measurement by many sensors. Satisfying this requirement alone makes the system place all sensors where conditions of measurement are least uncertain, leaving the system unprepared for sensing the stimuli that are useful but whose uncertainty is high. To prevent such gaps of allocation, we propose that minimal requirements should be twofold:

A. Reliability: Prefer low uncertainty.

B. Comprehensiveness: Measure all useful stimuli.

We formalize these requirements as follows. Let:

- $\Delta \in [a, b] \subset \mathbb{R}$ be the size of measuring device ("receptive field"),
- $U(\Delta) : \mathbb{R} \to \mathbb{R}$ be the uncertainty function associated with measuring devices of different size, and
- $r(\Delta) : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be the amount of resources allocated across Δ (Eq. the number of cells with receptive fields of size Δ).

Encouraging reliability. By requirement A, the system is penalized for allocating resources where uncertainty is high. This is achieved, for example, when the cost for placing resources at Δ is

$$k_1 U(\Delta) r(\Delta),$$

where k_1 is a positive constant. The higher the uncertainty at Δ, or the larger the amount of resources allocated to Δ, the higher the cost. Hence the total cost of allocation is:

$$J_1 = \int_a^b k_1 U(\Delta) r(x) d\Delta. \tag{S8}$$

Functional J_1 is minimal when all the detectors are allocated to (i.e., have the size of) Δ at the lowest value of $U(\Delta)$.

Encouraging comprehensiveness. By requirement B, the system is penalized for failing to measure particular stimuli. This is achieved, for example, when the allocation cost is

$$\frac{k_2}{r(\Delta)},$$

where k_2 is a positive constant. The total penalty of this type is:

$$J_2 = \int_a^b \frac{k_2}{r(x)} dx. \tag{S9}$$

Functional J_2 is large (infinite) when all resources are allocated to a small vicinity (one point). J_2 is small when $r(\Delta)$ are large for all Δ.

Prescription of allocation. The total penalty of requirements A and B is

$$J = \int_a^b k_1 U(\Delta) r(\Delta) + \frac{k_2}{r(\Delta)} d\Delta. \tag{S10}$$

Using standard tools of calculus of variations (e.g., Elsgolc, 2007) we find such function $r(\Delta)$ that minimizes J. In particular, we consider a variation of J with respect to changes of $r(\Delta)$:

$$\begin{aligned} \delta J = \int_a^b \frac{\partial}{\partial r(\Delta)} \left(k_1 U(\Delta) r(\Delta) + \frac{k_2}{r(\Delta)} \right) \delta r(\Delta) d\Delta \\ = \int_a^b \left(k_1 U(\Delta) - \frac{k_2}{r^2(\Delta)} \right) \delta r(\Delta) d\Delta. \end{aligned}$$

Because at optimal $r(\Delta)$ the value of δJ is zero for all $\delta r(\Delta)$, we deduce that conditions of optimality are:

$$U(\Delta) - \frac{k}{r^2(\Delta)} = 0, \; k = \frac{k_2}{k_1} \tag{S11}$$

In other words

$$r(\Delta) = \sqrt{\frac{k}{U(\Delta)}}. \tag{S12}$$

This $r(\Delta)$ is the prescription of optimal allocation.

Amount of resources. If the total amount or resources in the system is known and is C:

$$\int_a^b r(\Delta) = C, \tag{S13}$$

then we may modify coefficients k_1 and k_2 in Eq. S10, to make Eq. S10 consistent with Eq. S13. Or, we may use the method of Lagrange multipliers, looking for conditions where variation of the following functional vanishes:

$$\bar{J} = \int_a^b k_1 U(\Delta) r(\Delta) + \frac{k_2}{r(\Delta)} d\Delta + \lambda \left(\int_a^b r(\Delta) - C \right). \tag{S14}$$

We find Lagrange multiplier λ at which Eq. S13 is satisfied. The solution (using a method similar to that used for solving Eq. S11) is:

$$(k_1 U(\Delta) + \lambda) - \frac{k_2}{r^2(\Delta)} = 0 \;\Rightarrow r(\Delta) = \sqrt{\frac{k_2}{k_1 U(\Delta) + \lambda}} \tag{S15}$$

provided that

$$\int_a^b \sqrt{\frac{k_2}{k_1 U(\Delta) + \lambda}} d\Delta = C.$$

The latter constraint is used to find λ in Eq. S15. In either case, the shape of the optimal allocation function $r(\Delta)$ is determined by $U(\Delta)$, such that allocation function is maximal where $U(\Delta)$ is minimal. The formulation in Eq. S14 has an advantage. It allows one to derive optimal prescriptions under changes in the amount of resources allocated to the task, such as in selective attention.

Generalizations. In a multidimensional case, when Δ represents several variables (e.g., spatial and temporal extents of receptive fields, S and T), and $U(\cdot)$ is a function of many variables, the prescription is

$$r(s,t) = \sqrt{\frac{k}{U(s,t)}}.$$

Using the method of Lagrange multiplies, one can show that a similar result is obtained when the costs of reliability and comprehensibleness (Eqs. S8–S9) have more general formulations:

$$J_1 = \int_a^b k_1 U(\Delta) r^p(\Delta) d\Delta,$$
$$J_2 = \int_a^b k_2 \frac{1}{r^q(\Delta)} d\Delta, \;\; p, q \geq 1,$$

The previously derived prescription holds: allocate maximal amount of resources to conditions of minimal uncertainty.

References

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems.* Norwell, MA, USA: Kluwer Academic Publishers.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society (Supplement)*, *4*, 102–118.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory.* New York: John Wiley.

Elsgolc, L. D. (2007). *Calculus of variations.* Dover Publications. (Original work published in 1961.)

Gepshtein, S., Tyukin, I., & Albright, T. (2010). *The uncertainty principle of measurement in vision.* (Manuscript in preparation.)

Gepshtein, S., Tyukin, I., & Kubovy, M. (2007). The economics of motion perception and invariants of visual sensitivity. *Journal of Vision*, *7*(8), 1–18. (doi: 10.1167/7.8.8)

Landy, M., Maloney, L., Johnsten, E., & Young, M. (1995). Measurement and modeling of depth cue combinations: in defense of weak fusion. *Vision Research*, *35*, 389–412.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions.* New York: John Wiley.

Maloney, L. T., & Landy, M. S. (1989). Statistical framework for robust fusion of depth information. In W. A. Pearlman (Ed.), *Proc. spie vol. 1199, p. 1154-1163, visual communications and image processing iv, william a. pearlman; ed.* (p. 1154-1163).

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.

Taub, A. H. (Ed.). (1963). *John von Neumann: Collected works. Volume VI: Theory of games, astrophysics, hydrodynamics and meteorology.* New York, NY, USA: Pergamon Press.

von Neumann, J. (1928). Zur Theorie der Gesellschaftsspiele. [On the theory of games of strategy]. *Mathematische Annalen*, *100*, 295–320. (English translation in Taub, 1963.)

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge, UK: Cambridge University Press.